AD-A241 373

NAVAL POSTGRADUATE SCHOOL
Monterey, California

DTIC
ELECTE
OCT 10 1991
S B D

THESIS

ANALYSIS OF ENLISTMENT INCENTIVES
FOR HIGH QUALITY RECRUITS TO
THE UNITED STATES ARMY

by

Willis A. Woods

September, 1990

Thesis Co-Advisors: Laura Johnson
George Thomas

Approved for public release; distribution is unlimited.

91-12683

91 10 7 007

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Naval Postgraduate School

6b. OFFICE SYMBOL
OR

7a. NAME OF MONITORING ORGANIZATION
Naval Postgraduate School

6c. ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943-5000

7b. ADDRESS (City, State, and ZIP Code)
Monterey, California 93943-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION

8b. OFFICE SYMBOL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Including Security Classification)
ANALYSIS OF ENLISTMENT INCENTIVES FOR HIGH QUALITY RECRUITS TO THE UNITED STATES ARMY

12 PERSONAL AUTHOR(S)
WOODS, Willis Addison

13 TYPE OF REPORT
Master's Thesis

13b. TIME COVERED
FROM TO

14. DATE OF REPORT (Year, Month, Day)
1990, September

15. Page Count
86

16. SUPPLEMENTAL NOTATION
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Enlistment, Incentive, Motivation, New Recruit Survey

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
This thesis analyzes data from the 1988 New Recruit Survey (NRS) sponsored by the United States Army Recruiting Command to study incentives that motivate new recruits to enlist in the United States Army. Our purpose is to use discriminant analysis and logistic regression to identify those incentives that have the greatest effect on enlistees in the prime recruiting market and to compare the results of these two methods. We believe that the incentives identified will differ between high quality and non-high quality individuals where a high quality individual is defined as one who has a high school diploma and scores in categories I through IIIA on the Armed Forces Qualifiecation Test (AFQT). Demographic variables such as an individual's marital status and time spent in the labor force prior to enlisting in the Army were shown to influence enlistment incentives. Further, factor analysis of NRS responses identified four underlying factors which influenced recruits' enlistment motivations. However, these factors differed between racial groups and accurate models could only be developed for each racial group separately.

20 DISTRIBUTION/AVAILABILTIY OF ABSTRACT
X UNCLASSIFIED/UNLIMITED SAME AS RPT. DTIC

1a. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Laura Johnson

22b. TELEPHONE (Include Area Code)
(408)646-2569

22c. OFFICE SYMBOL
OR/Jo

DD Form 1473, JUN 86

Previous editions are obselete.
S/N 0102-LF-014-6603

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

Approved for public release; distribution is unlimited.

Analysis of Enlistment Incentives for High Quality Recruits to the United States Army

by

Willis A. Woods
Captain, United States Army
B.S., United States Military Academy, 1980

Submitted in paitial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS ANALYSIS
from the
NAVAL POSTGRADUATE SCHOOL
September 1990

Author: Willis A. Woods

Approved by: Laura Johnson, Co-Advisor

George Thomas, Co-Advisor

Linda Gorman, Second Reader

Peter Purdue, Chairman
Department of Operations Research

ABSTRACT

This thesis analyzes data from the 1988 New Recruit Survey (NRS) sponsored by the United States Army Recruiting Command to study incentives that motivate new recruits to enlist in the United States Army. Our purpose is to use discriminant analysis and logistic regression to identify those incentives that have the greatest effect on enlistees in the prime recruiting market and to compare the results of these two methods. We believe that the incentives identified will differ between high quality and non-high quality individuals where a high quality individual is defined as one who has a high school diploma and scores in categories I through IIIA on the Armed Forces Qualification Test (AFQT). Demographic variables such as an individual's marital status and time spent in the labor force prior to enlisting in the Army were shown to influence enlistment incentives. Further, factor analysis of NRS responses identified four underlying factors which influenced recruits' enlistment motivations. However, these factors differed between racial groups and accurate models could only be developed for each racial group separately.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

I. INTRODUCTION

The purpose of this thesis is twofold. First, to identify those enlistment incentives that have the greatest impact on high quality enlistees in the prime recruiting market. High quality recruits are individuals who score in categories I through IIIA on the Armed Forces Qualification Test (AFQT). Prime market recruits are considered to be 17 to 21 year old, male, high school diploma graduates. Second, this thesis will compare the results of two techniques for conducting the categorical data analysis supporting objective one described above. The two techniques that will be used are discriminant analysis and logistic regression analysis. The results of the analysis in this report will assist the U.S. Army Recruiting Command in developing advertising and compensation packages that will appeal to the demonstrated concerns of high quality enlistees in the prime recruiting market. Further, by identifying the expectations of recent enlistees, programs to fulfill these expectations and improve retention may be identified.

A. HIGH QUALITY AND PRIME MARKET

As the technical nature of military weapons systems continues to increase, the Army will also continue to depend on higher quality soldiers to maintain its effectiveness. The Chief of Staff of the Army General Carl E. Vuono states that

> In many conceivable contingencies potential adversaries throughout the world will enjoy numerical and geographical advantages, particularly in the early phases of a conflict. Those advantages demand that we have a high-quality force that, in turn, depends on quality people [Ref. 1:p. 12].

Regardless of these considerations, there is probably little argument that the military should be staffed by high quality soldiers. However, high quality, in terms of the needs of the Army should be carefully defined. According to a recent Department of the Army document,

> The affect [sic] of quality soldiers, defined as high school graduates who score in the top half of the Armed Forces Qualification Test (AFQT) (CAT I - IIIA), on individual and unit job performance is significant. Research conducted in 1989 has shown that excellent soldiers (CAT I-IIIA) performed 10 to 25 percent better than lower quality (CAT IV) soldiers in specific armor, infantry, artillery, and signal training tasks [Ref. 2:p. 18].

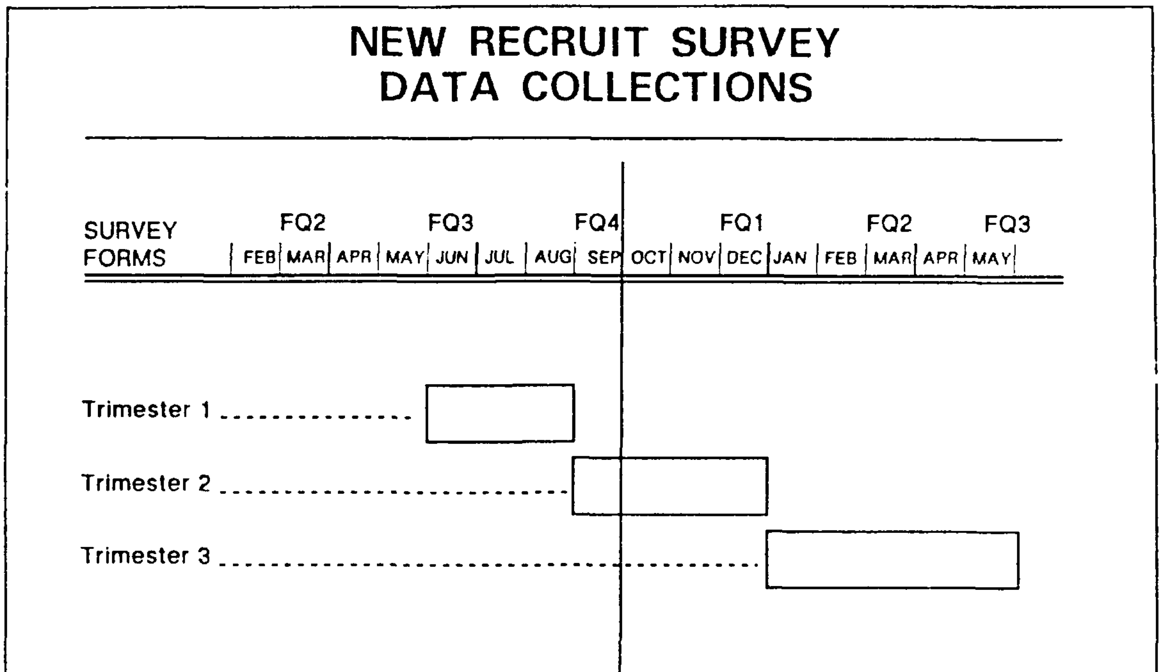
This indicates that there is good evidence in support of the definition of high quality stated above. Additionally, this study will restrict the high quality group to those recruits who graduated from high school with a diploma as opposed to a GED. Finally, the 17 through 21 year old entry age requirement is added to the high quality definition to identify the prime recruiting market. The concern is how to target these high quality, prime market individuals and provide incentives that will best attract them to join the Army.

B. TARGETING THE PRIME MARKET

Simply knowing what group of potential recruits the Army wants to attract is not enough. The Army must reach those potential recruits and convince them to join the Army. "Recruiting a quality force in the U.S. Army is predicated on adequate resources for advertising, incentive programs, and compensation..." [Ref. 2:p. 18]. Estimated Army advertising expenditures for the 1989 fiscal year are nearly $120 million [Ref. 3:p. 49]. To assist the Army in making the most effective use of these dollars, or perhaps even reduced resources, is a major concern of this study.

C. THE NEW RECRUIT SURVEY (NRS)

The Army's advertising agency, Young and Rubicam of New York City uses survey information from new recruits to determine how its advertising mission will be accomplished [Ref. 4:p. 19]. This thesis will use the 1988 edition of the same survey data which Young and Rubicam uses. These data come from the New Recruit Survey (NRS) which is sponsored by the United States Army Recruiting Command and prepared by the Data Recognition Corporation. The NRS is a "multi-year survey research endeavor...conducted to measure the enlistment motivations, attitudes, knowledge, and personal characteristics of new recruits at the time of their initial entry into the U.S. Army." The U.S. Army Research Institute (ARI) developed the NRS in 1982 under the direction of the Deputy Chief of Staff of the Army for Personnel. In 1984, the U.S. Army Recruiting Command (USAREC) assumed control of the NRS and until 1986 ARI maintained administration of the survey. After 1986, administration of the NRS was transferred to the Data Recognition Corporation and scheduled on a year-round basis [Ref. 5:p. ii]. Figure 1 shows the schedule for data collection for the data used in this thesis. These data provide survey responses from 5,863 new recruits of the active Army. Determining the best method of analyzing these data to study the impact of enlistment incentives on new recruits is a primary concern of this thesis.



Source: [Ref. 5:p. 3]

Figure 1 1988 New Recruit Survey Data Collections

II. LITERATURE REVIEW

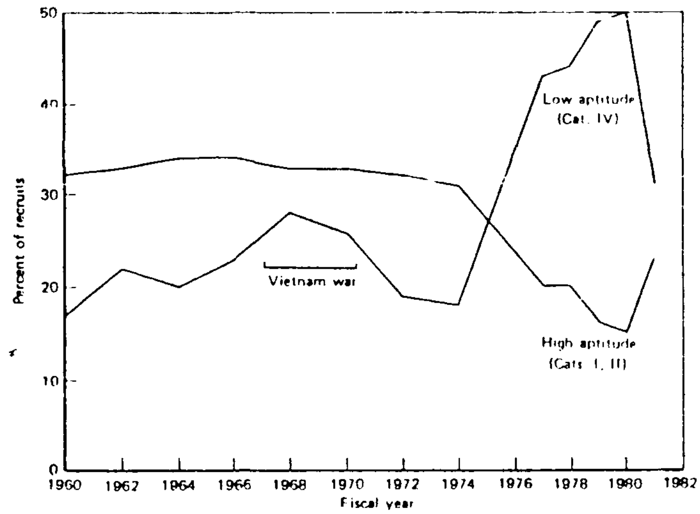
A. QUALITY, PERFORMANCE, AND ATTRITION

1. AFQT Scores

If there is any question that high quality soldiers (as defined in this thesis) perform better than low quality soldiers, despite the research supporting this statement, the military's inadvertent experiment of the late 1970's should provide a definitive answer.

> In 1980, the Department of Defense acknowledged that the aptitude battery used for determining enlistment eligibility between 1976 and 1980 had been "misnormed," which means that prospective recruits received higher scores than they would have received on a correctly calibrated test. As a result, many persons entered the services during the last half of the 1970's who did not meet draft-era enlistment standards; and in fact would not have been eligible to enlist with corrected scores [Ref. 6:p. 2].

The result of this calibration error was that by 1980 nearly fifty percent of all Army recruits were mental category IV, the lowest allowable level [Ref. 7:p. 1]. This result is shown in Figure 2.



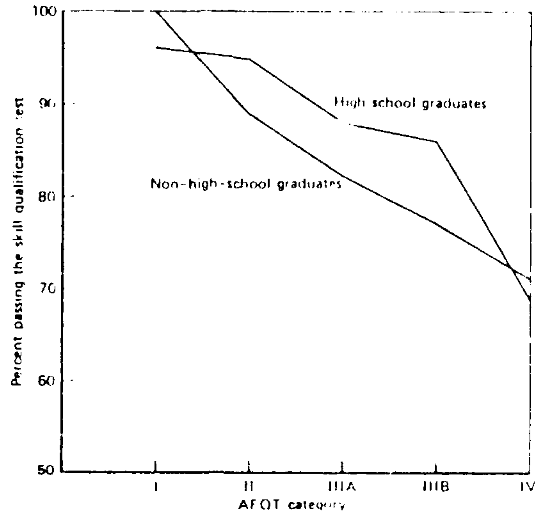
Source: [Ref. 7:p. 2]

Figure 2 Trends in high- and low-aptitude Army recruits

Results of the Army's Skill Qualification Tests (SQT), which are hands-on performance tests developed in the late 1970's for most Army jobs, can be used to assess the impact of this increase in low mental category recruits [Ref. 6:p. 6]. Figure 3 shows that "regardless of high school status, men in category IV (revised norms) are more likely to fail the minimum SQT standard than are persons in higher categories." [Ref. 7:p. 2]. The significance of these results was further amplified by

> Using two different types of on-the-job performance tests, and five different Army jobs, it has been shown that lower-aptitude recruits have significantly lower job-proficiency scores, and are significantly less likely to meet minimum proficiency standards than are higher-aptitude personnel. Therefore, the decline in ability standards in recent years has lowered Army manpower effectiveness by enlisting more personnel who are unable to meet minimum skill requirements [Ref. 6:p. 30].

These studies clearly indicate the need for high quality soldiers for the Army to maintain an acceptable level of performance.

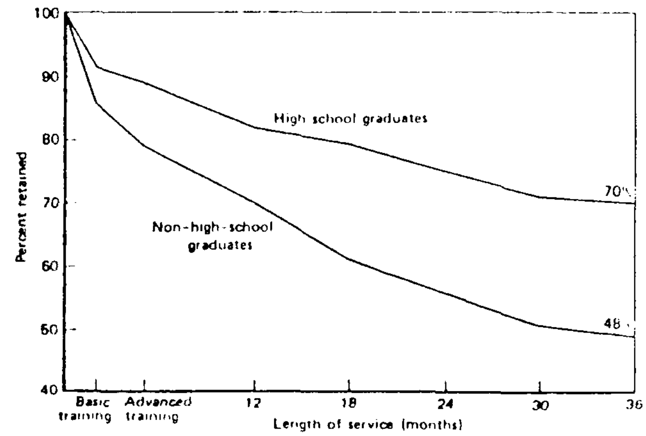


Source: [Ref. 7:p. 3]

Figure 3 Aptitude and Performance for Army Infantryman

2. High School Status

In spite of the poor job performance observed in low mental category soldiers, there is not a strong relationship between AFQT scores and attrition for first term recruits. There is, however, "a substantial association between high school status and attrition, both during and after training..." [Ref. 7:p. 6]. Figure 4 shows that 70% of high school graduates who enlist in the Infantry complete their initial term compared with only a 48% completion rate for non-high school graduates. Any soldier who fails to complete his initial enlistment represents a substantial lost investment for the Army. Therefore, it is critical that the Army attract recruits with the greatest probability of completing their enlistment. According to the research cited these people would be high school graduates.



Source: [Ref. 7:p. 7]

Figure 4 Army Infantrymen length of service and high school status

B. INCENTIVES, ADVERTISEMENT, AND ACCESSIONS

"Individuals choose to do something only if that choice makes them better off than other possible alternatives given their preferences and the information in their possession." [Ref. 8:p. 1]. This statement emphasizes the key to recruiting high quality soldiers and the principal issue of this thesis. To attract high quality recruits from the prime recruiting market, the Army must offer incentives that are important to these individuals. While this thesis will not specifically address advertising issues, potential recruits must receive information concerning enlistment incentives before the particular incentives will have any affect. Identifying those motivators that have attracted high quality recruits is critical in assisting the Army to develop incentives packages and advertising campaigns.

As stated earlier, we will use data from the 1988 New Recruit Survey (NRS). These data reflect the thoughts and opinions of only those individuals who enlisted in the Army. It should be acknowledged, that to best identify the motivators that attract

high quality recruits to join the Army, we would also like to have NRS data for those individuals who did not enlist in the Army. Unfortunately, data corresponding to the enlistment motivation questions used in this thesis are not currently available for individuals who have not enlisted in the Army.

III. DATA BASE AND METHODOLOGY

A. 1988 NEW RECRUIT SURVEY (NRS)

1. Survey Characteristics

The 1988 New Recruit Survey (NRS) was conducted in three trimesters as shown in Figure 1. The survey was administered at eight reception stations to a total of 5,863 U.S. Army active duty recruits as shown in Tables 1 and 2 below.[1]

TABLE 1 1988 NRS STATION SCHEDULE

Station	Weeks Survey Conducted at Station		
Ft. Benning	13 JUN 88	12 SEP 88	10 APR 89
Ft. Bliss	25 JUN 88	12 SEP 88	20 MAR 89
Ft. Dix	20 JUN 88	05 DEC 88	30 JAN 89
Ft. Jackson	27 JUN 88	07 NOV 88	20 FEB 89
Ft. Knox	01 AUG 88	24 OCT 88	27 FEB 89
Ft. Leonard Wood	18 JUL 88	14 NOV 88	15 MAY 89
Ft. McClellan	29 AUG 88	26 SEP 88	03 APR 89
Ft. Sill	08 AUG 88	17 OCT 88	23 JAN 89

[1]The complete survey also includes 2,242 Army National Guard and 1,626 Army Reserve recruits however this study is concerned only with active Army respondents.

TABLE 2 RESPONDENTS BY STATION

Station	Respondents
Ft. Benning	775
Ft. Bliss	292
Ft. Dix	947
Ft. Jackson	1196
Ft. Knox	731
Ft. Leonard Wood	847
Ft. McClellan	442
Ft. Sill	633

Selected tabulations of general characteristics of the 1988 NRS respondents are provided in Appendix A.

2. Enlistment Motivation Questions

The 1988 NRS contains 24 questions that specifically address the respondent's motivation to enlist in the Army. These 24 questions can be separated into two distinct groups.

The first group contains 22 questions that list a particular reason that could motivate a person to join the Army. The respondent is then asked to rate the importance of the stated reason for his decision to enlist. The possible responses are as follows:

- The reason was not at all important
- The reason was somewhat important
- The reason was very important

- I would not have enlisted except for this reason

The final 2 questions that deal with enlistment motivation each list ten reasons that could motivate a person to join the Army. Each respondent is asked to choose the one reason from this list of ten that was his *most important* reason for enlisting.

See Appendix B for a listing of these questions.

B. METHODOLOGY

This thesis will compare the results of discriminant analysis and logistic regression in identifying incentives that attract prime market recruits. The NRS survey data used are in SAS format and all data analysis and all techniques discussed will be implemented using SAS, Version 5.18.

1. Hypothesis

We hypothesize that the incentives which motivate prime market recruits to join the Army are different for high quality and non-high quality individuals. Two specific statistical techniques will be applied to the 1988 NRS data in order to identify the incentives providing the greatest motivation to high quality recruits in the prime market: discriminant analysis and logistic regression. The results of these techniques will be compared in relation to this hypothesis.

2. Discriminant Analysis

Procedure DISCRIM in SAS performs discriminant analysis which classifies observations into various groups based on a set of descriptive variables. This classification is accomplished by generating a set of functions whose coefficients are

chosen in a way such that the generalized squared distance between the variable values of an observation and the mean variable values of its assigned group is minimized [Ref. 9:p. 318]. The following discussion covers some of the theory behind discriminant analysis, and presents an example of the SAS DISCRIM procedure.

The example uses the 1988 NRS data and variables HIQUAL, T079, and T082 (these variables are chosen for purposes of this example only, and their selection has no other significance). Variable HIQUAL, the dependent variable, groups each observation as either *high quality* or *other* (according to the criteria developed earlier in the thesis). The variables T079 and T082 are used as the discriminating variables. These two questions ask the respondent to rate the importance of money for college (T079) and money for vo-tech school (T082) to their decision to enlist. A rating of one indicates that the reason was of no importance to the enlistees decision. A rating of four indicates that the respondent would not have enlisted except for that reason, and ratings of two or three indicate intermediate degrees of importance of that reason.

a. Generalized Squared Distance

The equation used by SAS for the generalized squared distance between an observation and its group mean is given in Equation 1 [Ref. 9:p. 318].

This equation is similar to the *Mahalanobis distance* which is the generalized squared distance between the mean variable values for each group.

b. SAS Output

The SAS DISCRIM procedure produces a set of linear discriminant functions. One function for each group in the analysis is included in the output. As stated above, the functions are generated such that the generalized squared distance

$$D_t^2(x) = (x-\bar{X}_t)^T \, COV^{-1} \, (x-\bar{X}_t)$$

where

$$\begin{aligned} D_t^2(x) &= \textit{generalized least squared distance} \\ &\quad \textit{from } x \textit{ to group } t \\ \bar{X}_t &= \textit{vector of means of variables for group } t \\ COV^{-1} &= \textit{inverse of pooled within groups} \\ &\quad \textit{covariance matrix} \end{aligned}$$

Equation 1 Generalized Least Square Distance

between an observation and its group mean is minimized. Equation 2 shows the general form of the discriminant functions.

$$Z_t = C_t + a_{1t} \, x_1 + \ldots + a_{nt} \, x_n$$

where

$$\begin{aligned} Z_t &= \textit{discriminant function for group } t \\ C_t &= \textit{constant term for group } t \\ a_{ik} &= \textit{coefficient for variable } i \textit{ group } k \\ x_{ik} &= \textit{value of variable} \end{aligned}$$

Equation 2 Discriminant Function

(1) Generating coefficients. The discriminant coefficients are based on the pooled within groups covariance matrix of the discriminating (independent) variables and the mean values for the discriminating variables for each group. Let $V=[v_{ij}]$ denote the covariance matrix as stated above then the matrix of coefficients $A=[a_{ij}]$ is given by: $A = V^{-1}\bar{X}$ [Ref. 10:p. 97]. Provided V is non-singular.

(2) Example. The DISCRIM procedure was used with variables HIQUAL, T079, and T082 as described above. The results of this procedure, listed in Equation 3, show the process of computing coefficients in this example.

POOLED WITHIN GROUPS COVARIANCE MATRIX

$$V = \begin{bmatrix} 0.9716 & 0.4439 \\ 0.4439 & 1.0493 \end{bmatrix}$$

INVERSE COVARIANCE MATRIX

$$V^{-1} = \begin{bmatrix} 1.2758 & -0.5397 \\ -0.5397 & 1.1813 \end{bmatrix}$$

GROUP MEANS

$$\bar{X} = \begin{bmatrix} 2.8665 & 2.4716 \\ 2.1035 & 2.265 \end{bmatrix}$$

COEFFICIENT MATRIX

$$A = U\bar{X} = \begin{bmatrix} 2.535 & 1.9308 \\ 0.9319 & 1.342 \end{bmatrix}$$

CONSTANTS

$$-C_j = \frac{\left(\sum_{l=1}^{m} \bar{x}_{lj}\, a_{lj}\right)}{2}$$

$$-c_1 = -4.63 \qquad -c_2 = -3.9$$

Equation 3 Deriving Discriminant Functions

From these results, the equations for the linear discriminant functions are shown in Equation 4.

$$Z_{hiqual} = -4.63 + 2.535(T079) + 0.9319(T082)$$

$$Z_{other} = -3.90 + 1.9308(T079) + 1.342(T082)$$

Equation 4 Example Discriminant Functions

These functions are used to classify observations in the respective groups by computing a score for each function based on the variable values for that observation and then classifying the observation to the group with the highest score.

(3) Computing one discriminant function. In the case of a model with only two groups, the two discriminant functions listed above can be directly converted to one equation. This is done by simply subtracting the coefficients for the second group from the coefficients for the first group which yields the single function shown in Equation 5 [Ref: 11:p. 260].

$$\begin{aligned} Z &= (a_{11} - a_{12})(T079) + (a_{21} - a_{22})(T082) \\ &= 0.6342(T079) + -0.4046(T082) \end{aligned}$$

Equation 5 One Discriminant Function

Note that the constant term is not included in this equation. Instead, a dividing point c is computed where $c = c_2 - c_1$ which results in a value of 0.8130 for this example. Note also the reverse order of subtraction to compute the dividing point. This is required since the constant term in the two discriminant functions is $-c_j$, not c_j (see Equation 3). Now this single function can be used to classify the observations as well. A score for each observation is computed using the function and the variable values for that observation. If the score is greater than the dividing point

c, then the observation is classified in group one if the score is less than the dividing point then it is classified in group two. The results are the same as the results obtained using two equations. [Ref. 11:p. 260]

c. Interpretation of Coefficients

The discriminant function coefficients indicate both the direction and degree of contribution each variable makes in classifying an observation. Consider the coefficients for the single discriminant function. A positive value for the coefficient indicates that observations with large values for the associated variable will tend to be classified in group one and visa versa. Further, these coefficients can be standardized by multiplying them by the pooled standard deviation for each variable. The magnitude of the standardized coefficient indicates the contribution of that variable to the discriminant function relative to the other coefficients. [Ref. 11:p. 257]

In the example, given a coefficient of +0.6342 for the variable T079 (which corresponds to money for college), a high score on this variable will contribute to that observation being classified as high quality. Or, in other words, a high quality individual will tend to be positively motivated to enlist in the Army given an incentive of earning money to attend college. On the other hand, the coefficient of -0.4046 for variable T082 (which corresponds to money for vo-tech school) indicates that the incentive of earning money to attend vocational or technical school provides the exact opposite effect. These results seem roughly logical but may not reflect the actual motivations of recruits. This could be due to the few number of variables used and the intentionally unsophisticated nature of the example model.

d. Posterior Probabilities

All of the previous discussions have considered only the discriminant function scores for a particular observation as a method of classifying the observation into a particular group. Another method of classification is by using the posterior probability of an observation belonging the assigned group [Ref. 11:p. 262]. The term posterior probability refers to the fact that the probability is computed *after* the analysis has been conducted. The posterior probability is the probability that an observation actually belongs to the group to which it was assigned during the discriminant analysis. This probability is also based on the generalized squared distance between the variable values of the observation and the mean variable values of the group to which it was assigned. Equation 6 lists the general formula for computing posterior probabilities [Ref. 11:p. 262].

$$p_t(x) = \frac{e^{-0.5 D_t^2(x)}}{\sum_{u=1}^{2} e^{-0.5 D_u^2(x)}}$$

where

$$t = group$$

$$D_t^2(x) = \text{generalized squared distance from } x \text{ to group } t$$

Equation 6 Posterior Probabilities

The posterior probabilities are particularly useful if one only wants to assign an observation to a group if it has a posterior probability above some threshold value. SAS uses the posterior probabilities to assign observations with the default threshold value of 0.5 (each observation assigned to the group with the greatest

posterior probability). The classification results using the default threshold value are the same as the previous two classification methods discussed.

3. Logistic Regression

Procedure LOGIST in SAS performs logistic regression to generate logistic function coefficients to classify observations into various groups based on a set of explanatory variables. The following discussion covers some of the theory behind logistic regression, how the logistic function coefficients are generated, and presents an example of the SAS LOGIST procedure.

The example uses the same variables as used in the discriminant analysis example so that direct comparisons may be made (again there is no significance to the particular explanatory variables used, they are for demonstration only). The example uses the 1988 NRS data and variables HIQUAL, T079, and T082. These are the same variables that were used in the example of discriminant analysis explained above.

a. SAS Output

(1) Developing the Logit Function. In this project, as in many social science scenarios, we are interested in predicting the group membership of a particular observation. In the case of a dichotomous response variable we can define group membership as follows:

$Y=1$ If the observation belongs to the first group

$Y=0$ If the observation belongs to the other group

Since the variable Y cannot assume continuous values, standard regression techniques are not appropriate. We can, however, use logistic regression to determine the probability that a particular observation belongs to a particular group based on the

values of the explanatory variables for that observation. The logistic equation used by SAS to predict the probability that Y=1 is shown in Equation 7 [Ref. 12:p. 270].

$$P_z = P[Y=1] = \frac{1}{1 + e^{-\alpha - X_i\beta}}$$

where

X_i = *the vector of variable values for the* i^{th} *observation*
β = *vector of regression parameters*
α = *the intercept parameter*

$$0 \le P_z \le 1$$

Equation 7 Logistic Function

Now we can also define the *odds* of belonging to group one as the probability of belonging to group one divided by the probability of not belonging to group one. This quantity is shown in Equation 8 [Ref. 11:p. 290].

$$odds = \frac{P_z}{1 - P_z}$$

$$0 \le odds \le \infty$$

Equation 8 Odds Function

Note the asymmetric range of both the logistic function and the odds function. By taking the natural logarithm of the odds function we can eliminate this asymmetry. This is known as the *logit* function and is illustrated in Equation 9 below. [Ref. 11:p. 290]

$$\begin{aligned} logit &= \ln\,(odds) \\ &= \ln\left(\frac{P_z}{1 - P_z}\right) \\ &= \alpha + x_i\beta \\ -\infty \le\ &logit \le \infty \end{aligned}$$

Equation 9 Logit Function

Note that the logit function is similar to the discriminant function in that the logit function is linear in the explanatory variables. The logit equation, however, has several attractive properties not found in the discriminant function that make it a good alternative for use in the analysis of categorical data.

> The fundamental assumption in logistic regression analysis is that ln(odds) is linearly related to the independent variables. No assumptions are made regarding the distributions of the X variables. In fact, one of the major advantages of this method is that the X variables may be discrete or continuous [Ref. 11:p. 291].

Discriminant analysis could be used to estimate the logistic parameters in Equation 9, but maximum likelihood estimates which depend only on the regression model should be used. Discriminant analysis requires multivariate normal explanatory variables while maximum likelihood estimates do not. In addition, logistic regression estimates are more robust than discriminant coefficient estimates. [Ref. 11:p. 291]

(2) Logistic Function Parameter Estimates. From the previous discussion we can define the probability that observation *i* belongs to a particular group as P_i. Then the relations in Equation 10 hold. [Ref. 13:p. 50]

$$P_i = P(Y_i = 1 \mid X_i)$$

$$1 - P_i = P(Y_i = 0 \mid X_i)$$

$$P(Y_i \mid X_i) = P_i^{Y_i} (1 - P_i)^{1-Y_i}$$

Equation 10 Probability of Y_i given X_i

From the equations above, the probability of observing a particular sample of N values of Y given all N sets of X_i observations is given by Equation 11. We define this as the likelihood function. [Ref. 13:p. 50]

$$L(Y \mid X,b) = P(Y \mid X) = \prod_{i=1}^{N} P_i^{Y_i}(1 - P_i)^{1-Y_i}$$

where b is the vector of regression coefficients

Equation 11 Likelihood Function

Now the maximum likelihood estimate for the vector of coefficients b, say β, is given by $L(Y|X) = \max_{b} L(Y|X,b)$. Since maximizing the natural logarithm of a function is equivalent to maximizing the function itself, we will take the natural logarithm of the likelihood function. Now we wish to maximize Equation 13 over b to find our estimates β. To accomplish this, we take the first derivative of Equation 13 with respect to each b in the coefficient vector and solve the resulting equation for zero. [Ref. 13:pp. 51-52]

$$\ln L(Y|X,b) = \sum_{i=1}^{N} \left[Y_i \ln P_i + (1-Y_i) \ln (1-P_i) \right]$$

Equation 13 Log-Likelihood Function

b. Example

(1) General. As stated earlier, the SAS LOGIST procedure was used with variables HIQUAL, T079, and T082 to illustrate a simple example of logistic regression. Omitting intermediate steps, the log likelihood function is given by Equation 14 where the subscripts 1 and 2 refer to variables T079 and T082 respectively.

$$\ln L(Y|X,b) = \sum_{i=1}^{N} \left[Y_i \ln P_i + (1-Y_i) \ln (1-P_i) \right]$$

where

$$P_i = \frac{1}{1 + e^{-(b_0 + b_1 x_{i1} + b_2 x_{i2})}}$$

Equation 14 Example Log-Likelihood Function

Now we take the first derivative of the log likelihood function with respect each b_i. set the resulting equations equal to zero and solve for the estimates β_i.

(2) Parameter Estimates. The parameter estimates generated in this example and the corresponding logit equation are shown in Equation 15.

This equation can be used to classify observations in a manner similar to that used in discriminant analysis. We compute the log odds for each observation using the logit equation and the explanatory variable values for that

$$\alpha = -0.493233$$
$$\beta_1 = 0.648140$$
$$\beta_2 = -0.431725$$

$$logit = \ln\left(\frac{P_z}{1 - P_z}\right)$$
$$= -0.493233 + 0.648140x_1 - 0.431725x_2$$

Equation 15 Example Logit Equation

observation. Since the range of the logit equation is symmetric about the origin, we simply assign the observation to group one (high quality) if the resulting value is greater than zero; or to group two (other) if the resulting value is less than zero.

c. Interpretation of Coefficients

The logistic function coefficients can be interpreted in the same manner as the discriminant function coefficients. They provide an indication of both the direction and degree of contribution for each variable to the classification [Ref. 11:p. 257]. A positive value for the coefficient indicates that observations with large values for the associated variable will tend to be classified to group one and visa versa.

For the example given a coefficient of +0.6481 for the variable T079 (which corresponds to money for college) means that a high score on this variable will contribute to that observation being classified as high quality. As observed in the discriminant model example, this indicates that a high quality individual will tend to be positively motivated to enlist in the Army based on the incentive of earning money to attend college. Again, as observed in the discriminant model example, the coefficient of -0.4317 for variable T082 (which corresponds to money for vo-tech school) indicates that the incentive of earning money to attend vocational or technical school

provides the exact opposite effect. These coefficients are also very similar in magnitude to those of the discriminant model example except for the dividing point. If we move the alpha term (-0.4932) in the logit equation to the left side of the equation (which of course changes the sign of the term yielding a value of +0.4932), this corresponds exactly to the discriminant model dividing point which had a value of +0.8130. Since individuals are assigned to the high quality group if the value of the assignment function used is greater than the dividing point, then in this example more individuals will be assigned to the high quality group when the discriminant model is used. This difference between the two models could be due to the assumptions required by the discriminant model. The accuracy of classification results for each model will be discussed later in the analysis portion of the study.

IV. ANALYSIS

A. OBJECTIVE

As mentioned previously, our primary objective is to identify enlistment incentives that motivate high quality recruits to enlist in the Army. Further, we want to contrast the results of discriminant analysis and logistic regression in identifying these incentives. To accomplish these objectives, we first developed models using both discriminant analysis and logistic regression to classify recruits as either high quality or other based on a set of explanatory variables. Additionally, we analyzed the models to determine the relative importance of each explanatory variable in the classification of high quality recruits to identify factors providing the greatest enlistment incentives to this group. Lastly, a comparison of the results from the two models was done.

B. EXPLANATORY VARIABLES

1. Variables Initially Selected

In choosing variables to use in the analysis, two primary considerations were addressed. First, we believe that certain strictly demographic factors will affect one's motivation to enlist in the Army. Of the NRS variables available in this category, we felt that race, marital status, additional education since high school, and potential time in the job market were variables that would most influence enlistment motivation.

Of the 1988 NRS respondents, 99% had no additional education since high school, so this variable was dropped from consideration. For race, less than 4% of the respondents were listed as Indian/Alaskan or Asian/Pacific so race was transformed

into a dichotomous variable where the possible responses were White/Other or Black. The Indian/Alaskan and Asian/Pacific categories were added to the White/Other response because of the relatively small increase this causes to the White/Other group. Additionally, only 0.1% of the enlistees responded, that they were divorced to the marital status question so these responses were added to the single category to produce another dichotomous variable having the possible responses of either married or single.

The variable to measure potential time in the job market deserves special explanation. We believe that potential time in the job market can be approximated by the time between the date that the respondent graduated from high school and the date that he took the NRS survey. During this time the individual can be considered in the job market. Although we really have no knowledge of whether the person actually was working, or seeking work, we believe that this period may contribute to the individual's enlistment motivations. Both the high school graduation date and the survey date are available in the NRS data base, so by simply subtracting one from the other, we arrive at the number of months that the person was potentially in the job market. This variable was divided into two levels where the first level is less than or equal to one year, and the second level is greater than one year in the potential job market.

The second consideration is the stated reasons for enlisting in the Army as possible explanatory variables. Questions in the survey that will be most indicative of enlistment motivations are the twenty-two weighted response questions that specifically address reasons for enlisting. In these questions, the respondent is presented with a particular reason for enlisting in the Army and he must rate the

importance of this reason toward his decision to enlist. An answer of "1" indicates that the reason was of no importance, "2" indicates fairly important, "3" indicates very important, and an answer of "4" indicates that the respondent would not have joined except for this reason.

While we believe that these questions can be used as good indicators of motivation to enlist in the Army, we do not believe that these variables can be considered independent. In order to deal with the dependence between these variables and to also reduce the number of explanatory variables in our analysis, we used principal factor analysis to identify the relationships between these variables and to help develop new orthogonal variables to use in our analytical models. The results of this factor analysis are covered in the next section.

2. Factor Analysis

a. General

As discussed in the previous section, we believe that the twenty-two weighted response variables in the NRS may be good predictors of enlistment motivators, but we also believe that they are correlated with each other. To limit this dependence among the variables, we used factor analysis to develop a new set of variables which have a minimum of correlation with each other. The basic idea behind factor analysis is that the original set of variables can be described by a smaller underlying set of factors. Factor analysis is a formal method of determining how many of these underlying factors exist and the weight that each of the original variables contributes to the individual factors [Ref. 10:p. 9]. In effect, the smaller set of underlying factors becomes a linear combination of the original variables.

b. Questions With Greatest Loadings per Factor

The factor analysis procedure identified four factors underlying the twenty-two weighted response variables. The factors can be subjectively named by observing which reasons are weighted most heavily in the rotated factor pattern. The four factors with their subjective names and most heavily weighted variables are listed below. See Appendix C for a complete table of factor loadings.

(1) Factor 1 - Better myself

• Importance of becoming a responsible person	0.72
• Importance of becoming more self-reliant	0.69
• Importance of becoming a better individual	0.66
• Importance of a chance to better myself	0.51
• Importance of money for college	0.24

(2) Factor 2 - Serve my country/be a leader

• Importance of wanting to be a soldier	0.67
• Importance of serving my country	0.64
• Importance of leadership training	0.49
• Importance of physical training	0.46
• Importance of proving I can make it	0.34
• Importance of family tradition to serve	0.31

(3) Factor 3 - Money/benefits/job

• Importance of fringe benefits	0.58
• Importance of retirement benefits	0.53
• Importance of getting a better job	0.51
• Importance of skill training	0.43
• Importance of earning more money	0.41
• Importance of money for vo-tech school	0.29
• Importance of unemployment	0.23

(4) Factor 4 - Get away from home/travel

• Importance of being away from home	0.43
• Importance of time to decide life plans	0.39
• Importance of escaping personal problem	0.36
• Importance of travel	0.29

3. Final Variables Selected

Based on the subjective beliefs concerning demographic variables and the factor analysis mentioned above, the following variables were selected for inclusion in our models:

- Race
- Marital Status
- Potential Experience in the Labor Force
- Factor 1 (Better myself)

- Factor 2 (Serve my country/be a leader)
- Factor 3 (Money/benefits/job)
- Factor 4 (Get away from home/travel)

C. RESULTS OF COMBINED MODELS

The initial models developed with the variables listed above are termed "combined models" because both racial categories were included in the factor analysis and in the two models. Some of the results discussed below indicate that this may not be the best procedure to use and alternative methods, with results, are also presented.

1. Discriminant Analysis Model

a. Classification Equations

Using the variables described above, the SAS procedure DISCRIM was used to conduct a discriminant analysis between the high quality and non-high quality survey respondents. The standard procedure output is one classification equation for each group. Observations are then assigned to the group on which they have the highest score based on these classification equations. The two equations are listed below.

$$
\begin{array}{ll}
Z_{high} = -0.24 & Z_{other} = -0.77 \\
\quad 0.81*(Race) & \quad 2.58*(Race) \\
\quad 0.67*(Marital\ Status) & \quad 1.11*(Marital\ Status) \\
\quad 1.27*(Labor\ Force) & \quad 1.20*(Labor\ Force) \\
\quad -0.09*(Factor\ 1) & \quad -0.20*(Factor\ 1) \\
\quad -0.04*(Factor\ 2) & \quad 0.20*(Factor\ 2) \\
\quad -0.18*(Factor\ 3) & \quad 0.11*(Factor\ 3) \\
\quad 0.09*(Factor\ 4) & \quad 0.07*(Factor\ 4)
\end{array}
$$

Equation 16 Discriminant Classification Equations

b. Classification Results

Based on the equations above, the classification results (using the same data as the coefficients were generated from) are shown in Table 3 below.

TABLE 3 DISCRIMINANT MODEL CLASSIFICATION RESULTS

Actual Group	**Classified As Group**	
	High	*Other*
High	**1256 (80.67%)**	**301 (19.33%)**
Other	**534 (50.71%)**	**519 (49.29%)**

2. Logistic Regression Model

a. Classification Equation

The SAS procedure LOGIST was used to perform logistic regression using the quality variable as the response variable and the dependent variables described above as the explanatory variables. The model and coefficients generated are shown in Equation 17 below.

b. Classification Results

Based on Equation 17, the classification results are shown in Table 4 below.

$$P[\mathit{Quality}=\mathit{High}] = \frac{1}{1+e^{-\alpha-\beta X}}$$

where

$$\alpha = 0.84$$

and

$$\beta\ X = \begin{cases} -1.60*(\mathit{Race}) \\ -0.41*(\mathit{Marital\ Status}) \\ \ \ 0.08*(\mathit{Labor\ Force}) \\ \ \ 0.12*(\mathit{Factor}\ 1) \\ -0.24*(\mathit{Factor}\ 2) \\ -0.30*(\mathit{Factor}\ 3) \\ \ \ 0.02*(\mathit{Factor}\ 4) \end{cases}$$

Equation 17 Logistic Classification Equation

TABLE 4 LOGISTIC MODEL CLASSIFICATION RESULTS

Actual Group	**Classified As Group**	
	High	*Other*
High	**1323 (85.00%)**	**234 (15.00%)**
Other	**586 (55.70%)**	**467 (44.30%)**

3. Comparison of The Two Methods

a. Theoretical

As mentioned earlier, the discriminant classification equation and the logit function are both linear in the explanatory variables. Additionally, except when model assumptions are violated, we would expect results from the two procedures to be quite similar. If the explanatory variables are multivariate normal (as assumed by the discriminant model), then the same level of precision as with logistic regression can be achieved even when a smaller sample size is used [Ref. 11:p. 291]. However, "the estimates of the coefficients or the probabilities derived from the two methods

will rarely be substantially different from each other, whether or not the multivariate normality assumption is satisfied" [Ref. 11:p. 291].

b. Observed

As expected the results indicate that the two methods are fairly similar in classifying respondents. Although, the discriminant procedure is better at classifying the other category of respondents and the logistic procedure is better at classifying the high quality respondents, these differences are fairly small. Further, the previous results only allow us to compare the two methods based on their relative classification results. We can, however arrive at a more direct comparison of the two classification methods with some simple manipulation of the respective classification equations.

By subtracting the corresponding coefficients of the two classification equations from the discriminant analysis, we can generate a single classification equation. Further by subtracting the constant terms from these two equations, we find a "dividing point" for our equation. Now by evaluating an observation on this new equation, we can classify the individual depending on whether the resulting value of the equation is greater than or less than the dividing point.

Similarly, if we use the "log odds" form of the logistic regression equation, we have an equation of the same form as the single discriminant equation above. In fact, the discriminant coefficients could have been used in the logistic model in the first place but using maximum likelihood estimates instead allows us to avoid the multivariate normal requirements of discriminant analysis. These new equations are shown in Equation 18 below.

Discriminant classify high if Z > -0.53 where	*Logistic classify high if Z > -0.84 where*
$Z = -1.77*(Race)$	$Z = -1.60*(Race)$
$-0.43*(Marital\ Status)$	$-0.41*(Marital\ Status)$
$0.08*(Labor\ Force)$	$0.08*(Labor\ Force)$
$0.12*(Factor\ 1)$	$0.12*(Factor\ 1)$
$-0.24*(Factor\ 2)$	$-0.24*(Factor\ 2)$
$-0.29*(Factor\ 3)$	$-0.30*(Factor\ 3)$
$0.02*(Factor\ 4)$	$0.02*(Factor\ 4)$

Equation 18 Comparison of Classification Equations

These equations indicate that potential labor force experience, Factor 1, and Factor 4 are important in determining if a respondent is classified as high quality. This gives some indication that high quality enlistees spent more potential time in the labor force prior to joining the Army. Additionally, high quality respondents were more interested in becoming better, more responsible people and having an opportunity to travel, as indicated by Factor 1 and Factor 4 respectively. Conversely, Marital Status, Factor 2, and Factor 3 all have negative coefficients indicating that these variables do not contribute to classifying individuals as high quality (note that the race variable has not been mentioned here, the section below will explain why). This indicates that if a high quality individual is married he may be less inclined to join the Army. Also, the negative coefficient associated with Factor 2 indicates the high quality recruits were less likely to be motivated by a desire to serve when they enlisted in the Army. Similarly, the negative coefficient for Factor 3 indicates that high quality recruits are less interested in incentives directly associated with monetary compensation, getting a job, or future benefits such as retirement.

Unfortunately, neither the labor force variable nor Factor 4 are significant at the 0.05 level (all other variables are significant at this level) based on the logistic regression model. Naturally, this leads to some skepticism regarding any conclusions drawn based on these variables despite the relatively accurate classification results.

4. Problems

While the preceding results appear to be encouraging, a closer analysis indicates that both discriminant analysis and logistic regression are poor at classifying black respondents. Tables 5 through 8 below indicate the classification results for each procedure by racial category.

TABLE 5 DISCRIMINANT MODEL (WHITE ONLY)

Actual Group	**Classified As Group**	
	High	*Other*
High	**1256 (94.15%)**	**78 (05.85%)**
Other	**534 (90.36%)**	**57 (09.64%)**

TABLE 6 LOGISTIC MODEL (WHITE ONLY)

Actual Group	**Classified As Group**	
	High	*Other*
High	**1322 (99.10%)**	**12 (00.90%)**
Other	**584 (98.82%)**	**7 (01.18%)**

TABLE 7 DISCRIMINANT MODEL (BLACK ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**0 (00.00%)**	**229 (100.00%)**
Other	**0 (00.00%)**	**467 (100.00%)**

TABLE 8 LOGISTIC MODEL (BLACK ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**1 (00.44%)**	**228 (99.56%)**
Other	**2 (00.43%)**	**465 (99.57%)**

Clearly, the "combined" models do not accurately model quality according to the two racial categories. We believe that this may be due to sociological differences which influence incentives that may vary between the two racial groups. Since the sample population is mainly (73.4%) white, the sociological characteristics of black respondents could be misrepresented in the factor analysis procedure. To attempt to correct this deficiency, we replicated the previous work separately for each racial group. The results of these "separated" models are presented in the next section.

D. RESULTS OF MODELS FOR BLACK GROUP ONLY

1. Factor Analysis

The factor analysis procedure for the black only racial group again identified four factors underlying the twenty-two weighted response variables. The first three factors are close to the first three factors in the combined factor analysis, however, the

fourth factor doesn't appear to follow a single distinct pattern. The factors can be subjectively named by observing which reasons are weighted most heavily in the rotated factor pattern. The four factors with their subjective names and most heavily weighted variables are listed below (the subscript "B" is added to the factor number to indicate that the factors were derived from the black respondents only). See Appendix C for a complete table of factor loadings.

a. Factor 1_B - Better myself

• Importance of becoming a responsible person	0.70
• Importance of becoming more self-reliant	0.63
• Importance of becoming a better individual	0.59
• Importance of a chance to better myself	0.55

b. Factor 2_B - Serve my country/be a leader

• Importance of wanting to be a soldier	0.70
• Importance of serving my country	0.67
• Importance of leadership training	0.50
• Importance of physical training	0.45
• Importance of travel	0.23

c. Factor 3_B - Money/benefits/job

• Importance of fringe benefits	0.57
• Importance of retirement benefits	0.54
• Importance of getting a better job	0.44

- Importance of money for vo-tech school 0.44
- Importance of skill training 0.44
- Importance of earning more money 0.43
- Importance of money for college 0.33

d. Factor 4_B - Other

- Importance of time to decide life plans 0.46
- Importance of being away from home 0.43
- Importance of escaping a personal problem 0.42
- Importance of unemployment 0.39
- Importance of family tradition to serve 0.35
- Importance of proving I can make it 0.33

2. Discriminant Analysis Model

a. Classification Equations

Using the variables described above, the SAS procedure DISCRIM was used to conduct a discriminant analysis between the high quality and non-high quality survey respondents. The standard procedure output is one classification equation for each group. Observations are then assigned to the group on which they have the highest score based on these classification equations. The two equations are listed below.

$$
\begin{array}{ll}
Z_{high} = -0.17 & Z_{other} = -0.26 \\
\quad 0.38 * (\mathit{Marital\ Status}) & \quad 0.73 * (\mathit{Marital\ Status}) \\
\quad 1.13 * (\mathit{Labor\ Force}) & \quad 1.55 * (\mathit{Labor\ Force}) \\
\quad -0.01 * (\mathit{Factor}\ 1_B) & \quad -0.05 * (\mathit{Factor}\ 1_B) \\
\quad -0.28 * (\mathit{Factor}\ 2_B) & \quad 0.11 * (\mathit{Factor}\ 2_B) \\
\quad -0.08 * (\mathit{Factor}\ 3_B) & \quad -0.12 * (\mathit{Factor}\ 3_B) \\
\quad -0.11 * (\mathit{Factor}\ 4_B) & \quad 0.13 * (\mathit{Factor}\ 4_B)
\end{array}
$$

Equation 19 Discriminant Model (Black Only)

b. Classification Results

Based on the equations above, the classification results are in Table 9 below.

TABLE 9 CLASSIFICATION RESULTS (BLACK ONLY)

Actual Group	**Classified As Group**	
	High	*Other*
High	**129 (60.56%)**	**84 (39.44%)**
Other	**201 (43.89%)**	**257 (56.11%)**

3. Logistic Regression Model

a. Classification Equation

The SAS procedure LOGIST was used to perform logistic regression using the quality variable as the response variable and the dependent variables described above as the explanatory variables. The model and coefficients generated are shown in Equation 20 below.

b. Classification Results

The classification results are in Table 10 below.

$$P[Quality{=}High] = \frac{1}{1+e^{-\alpha-\beta X}}$$

where

$$\alpha = -0.67$$

and

$$\beta X = \begin{cases} -0.41*(Marital\ Status) \\ -0.43*(Labor\ Force) \\ \ \ 0.05*(Factor\ 1_B) \\ -0.40*(Factor\ 2_B) \\ \ \ 0.04*(Factor\ 3_B) \\ -0.27*(Factor\ 4_B) \end{cases}$$

Equation 20 Logistic Model (Black Only)

TABLE 10 CLASSIFICATION RESULTS (BLACK ONLY)

Actual Group	**Classified As Group**	
	High	*Other*
High	**3 (01.00%)**	**210 (99.00%)**
Other	**9 (02.00%)**	**449 (98.00%)**

4. Comparison of The Two Methods

The previous results only allow us to compare the two methods based on their relative classification results. We can, however arrive at a more direct comparison of the two classification methods with some simple manipulation of the respective classification equations.

The procedures to generate these equations were described earlier and are not repeated here. The new equations are shown in Equation 21 below.

Discriminant	*Logistic*
classify high	*classify high*
if Z > -0.098 where	*if Z > 0.666 where*
$Z = -0.34*(\textit{Marital Status})$	$Z = -0.41*(\textit{Marital Status})$
$-0.42*(\textit{Labor Force})$	$-0.43*(\textit{Labor Force})$
$0.04*(\textit{Factor } 1_B)$	$0.05*(\textit{Factor } 1_B)$
$-0.39*(\textit{Factor } 2_B)$	$-0.40*(\textit{Factor } 2_B)$
$0.04*(\textit{Factor } 3_B)$	$0.04*(\textit{Factor } 3_B)$
$-0.24*(\textit{Factor } 4_B)$	$-0.27*(\textit{Factor } 4_B)$

Equation 21 Comparison of Classification Equations (Black Only)

These equations indicate that the reason that the logistic equation is so poor at correctly classifying high quality respondents is because of the unusually high intercept term. Later, we will present a technique to compensate for this fact and improve the classification results for the logistic model.

E. RESULTS OF MODELS FOR WHITE GROUP ONLY

1. Factor Analysis

The factor analysis procedure for the white only racial group again identified four factors underlying the twenty-two weighted response variables. All four factors are close to the factors identified in the combined factor analysis. The factors can be subjectively named by observing which reasons are weighted most heavily in the rotated factor pattern. The four factors with their subjective names and most heavily weighted variables are listed below (the subscript "W" is added to the factor number to indicate that the factors were derived from the white respondents only). See Appendix C for a complete table of factor loadings.

a. Factor 1_W - Better myself

• Importance of becoming a responsible person	0.77
• Importance of becoming a better individual	0.75
• Importance of becoming more self-reliant	0.72
• Importance of a chance to better myself	0.54
• Importance of leadership training	0.48
• Importance of physical training	0.43

b. Factor 2_W - Serve my country/be a soldier

• Importance of wanting to be a soldier	0.61
• Importance of serving my country	0.57
• Importance of proving I can make it	0.36
• Importance of family tradition to serve	0.31

c. Factor 3_W - Benefits/job

• Importance of fringe benefits	0.56
• Importance of getting a better job	0.53
• Importance of retirement benefits	0.49
• Importance of skill training	0.45
• Importance of earning more money	0.42
• Importance of unemployment	0.30

d. Factor 4_W - Travel/education

• Importance of money for college	0.44
• Importance of money for vo/tech school	0.39
• Importance of time to decide life plans	0.39
• Importance of being away from home	0.38
• Importance of travel	0.34
• Importance of escaping a personal problem	0.24

2. Discriminant Analysis Model

a. Classification Equations

As with the Black only model, the SAS procedure DISCRIM was used to conduct a discriminant analysis between the high quality and other survey respondents. Again, observations are assigned to the group on which they have the highest score based on these classification equations. The two equations are listed in Equation 22 below.

$$
\begin{array}{ll}
Z_{high} = -0.18 & Z_{other} = -0.19 \\
\quad 0.66 * (\mathit{Marital\ Status}) & \quad 0.66 * (\mathit{Marital\ Status}) \\
\quad 1.29 * (\mathit{Labor\ Force}) & \quad 1.21 * (\mathit{Labor\ Force}) \\
\quad -0.06 * (\mathit{Factor\ } 1_W) & \quad 0.01 * (\mathit{Factor\ } 1_W) \\
\quad -0.04 * (\mathit{Factor\ } 2_W) & \quad 0.18 * (\mathit{Factor\ } 2_W) \\
\quad -1.19 * (\mathit{Factor\ } 3_W) & \quad 0.22 * (\mathit{Factor\ } 3_W) \\
\quad 0.16 * (\mathit{Factor\ } 4_W) & \quad -0.10 * (\mathit{Factor\ } 4_W)
\end{array}
$$

Equation 22 Discriminant Model (White Only)

b. Classification Results

Based on the equations above, the classification results are in Table 11 below.

TABLE 11 CLASSIFICATION RESULTS (WHITE ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**763 (57.72%)**	**559 (42.28%)**
Other	**250 (41.74%)**	**349 (58.26%)**

3. Logistic Regression Model

a. Classification Equation

As before, the SAS procedure LOGIST was used to perform logistic regression. The model and coefficients generated are shown in Equation 23 below.

$$P[\mathit{Quality}=\mathit{High}] = \frac{1}{1+e^{-\alpha-\beta X}}$$

where

$$\alpha = 0.80$$

and

$$\beta\ X = \begin{cases} 0.01 * (\mathit{Marital\ Status}) \\ 0.08 * (\mathit{Labor\ Force}) \\ -0.07 * (\mathit{Factor}\ 1_w) \\ -0.21 * (\mathit{Factor}\ 2_w) \\ -0.40 * (\mathit{Factor}\ 3_w) \\ 0.26 * (\mathit{Factor}\ 4_w) \end{cases}$$

Equation 23 Logistic Model (White Only)

b. Classification Results

Based on the equation above, the classification results are in Table 12 below.

TABLE 12 CLASSIFICATION RESULTS (WHITE ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**1303 (99.00%)**	**19 (01.00%)**
Other	**584 (97.00%)**	**15 (03.00%)**

4. Comparison Of The Two Methods

As discussed in the Black only analysis section, we make some simple manipulations of the above classification equations to arrive at a more direct comparison of the two classification methods. The results of this process are shown in Equation 24 below.

Discriminant	*Logistic*
classify high	*classify high*
if $Z > -0.011$ *where*	*if* $Z > -0.802$ *where*
$Z = -0.00*(\textit{Marital Status})$	$Z = 0.01*(\textit{Marital Status})$
$0.08*(\textit{Labor Force})$	$0.08*(\textit{Labor Force})$
$-0.07*(\textit{Factor } 1_W)$	$-0.07*(\textit{Factor } 1_W)$
$-0.21*(\textit{Factor } 2_W)$	$-0.21*(\textit{Factor } 2_W)$
$-0.40*(\textit{Factor } 3_W)$	$-0.40*(\textit{Factor } 3_W)$
$0.26*(\textit{Factor } 4_W)$	$0.26*(\textit{Factor } 4_W)$

Equation 24 Comparison of Classification Equations (White Only)

Just as in the Black only analysis, we observe that logistic regression poorly classifies high quality respondents because of the unusually high intercept term. The

next section presents a technique to compensate for this fact and improve the classification results for the logistic model.

In contrast to the Black only equations, the coefficients associated with the Marital Status, Labor Force, Factor 1, Factor 3, and Factor 4 variables are of opposite sign in the White only equations. This indicates that these variables have exactly the opposite effect on high quality individuals based on their race. Recall, however, that the respective factor variables are not identical for each racial category and as such cannot be directly compared. These results and their interpretation for each racial category will be discussed further in the conclusion section of the thesis.

F. ADJUSTED LOGISTIC MODEL

1. General

The results of the previous section show that modeling the data separately by each race improves the classification results for the discriminant models but not for the logistic models.

Recall that the logistic model is merely a probability of group membership. Each observation is assigned as high quality if it has greater that a 0.5 probability of being in that group based on the explanatory variables; otherwise, the observation is assigned to the other group. We may however, specify a different threshold probability in order to attempt to correct the poor results of the logistic model.

2. Adjusted Assignment Probability

As discussed earlier, we can influence the classification procedure in the logistic model by adjusting the probability threshold level for group assignment. For the Black race category, the default threshold of 0.5 was shown to assign 98% of the

respondents to the other group when only 68% of the respondents were actually in this group. For the White race category, 98% of the respondents were assigned to the high quality group when only about 69% of the respondents were actually in the high quality group.

These classification results indicate that the threshold probability for the Black race category is too high and that the threshold probability for the White race category is too low. Assuming that we would desire the classification results to be similar to those for the discriminant models, we can experiment with different threshold probabilities to accomplish this goal.

For the Black race category, a threshold probability of 0.325 results in the classification results shown in Table 13. For the White race category, a threshold probability of 0.685 results in the classification results in Table 14.

TABLE 13 CLASSIFICATION RESULTS (BLACK ONLY, p=0.325)

Actual Group	**Classified As Group**	
	High	*Other*
High	**124 (58.21%)**	**89 (41.78%)**
Other	**196 (42.79%)**	**262 (57.21%)**

TABLE 14 CLASSIFICATION RESULTS (WHITE ONLY, p=0.685)

Actual Group	**Classified As Group**	
	High	*Other*
High	**778 (58.85%)**	**544 (41.15%)**
Other	**259 (43.24%)**	**340 (56.76%)**

These results are much closer to the results found in the discriminant models for the respective race categories and provide much more balanced correct classifications between the high quality and other groups.

G. CLASSIFICATION RESULTS USING DIFFERENT DATA

As mentioned before, all classification results reported earlier in the thesis are computed by experimental classification of the data that was used to generate the model coefficients. This data represents only 80% of the entire sample of respondents. The other 20% of the sample data points were withheld in order to provide another sample to check the models. The classifications listed in the main analysis of the thesis were repeated using this smaller data set, and the results were quite similar to those using the larger data set. All small sample classification tables are listed in Appendix D.

V. CONCLUSIONS

Our objective has been to identify those enlistment incentives that have the greatest impact on enlistees in the prime recruiting market. We hypothesized that the incentives which motivate prime market recruits to join the Army are different for high quality and non-high quality individuals. Further, we wanted to compare the results of discriminant analysis and logistic regression in conducting the categorical data analysis to identify these enlistment incentives. We have been able to identify enlistment incentives as desired, however, our results indicate models based on either technique should be developed separately for each racial group under consideration. Further, our analysis indicates that there may be certain conditions that cause the use of one model over the other to be preferable.

A. COMBINED MODELS

Due to the poor classification results for Black respondents observed in the "combined" models discussed in the previous chapter, these models are considered to be of limited value in correctly identifying enlistment incentives. However, the results of the "combined" models do provide some indication that the discriminant analysis and logistic regression models provide comparable classification results.

B. SEPARATED MODELS

Since the classification results for each racial category were so poor, we believe that models, separated by race, are required to accurately identify incentives important to all sample respondents.

1. Factor Analysis

As expected, conducting factor analysis separately for each racial category identified different factors for blacks and whites. Although these differences are not dramatic, they confirm the belief that separate models for each race are required.

2. Model Effectiveness

The "separated" discriminant analysis models are fairly successful in predicting quality group membership for both the Black and the White racial groups. Therefore, these models can be effectively used to identify incentives for the high quality respondents.

The "separated" logistic regression models are highly inaccurate in group classification at the 0.5 threshold probability and must be modified in order to obtain acceptable classification results. By adjusting the threshold classification probabilities *more accurate classification results can be achieved.*

3. Important Enlistment Incentives

a. Black Racial Category

For respondents in the Black racial category, both models identified Factor 1_B and Factor 3_B as explanatory variables contributing to high category classification. This first factor indicates that the black, high quality enlistees are concerned with becoming more responsible, more self-reliant people. Additionally, the second factor, indicates that the black, high quality enlistees are concerned with earning money and receiving benefits in the Army. The second factor also includes such concerns as receiving skill training directly and in receiving money to use for education at vo-tech schools or college. According to the "separated" models these

incentives were most influential in attracting the high quality black enlistees surveyed to enlist in the Army, and these incentives may be effective in attracting future enlistees to join the Army.

b. White Racial Category

For respondents in the White racial category, both models identified potential labor force experience and Factor 4_W as explanatory variables contributing to high category classification. While the potential labor force experience variable does not specifically address enlistment incentives, this indicates that for this sample of recruits, white respondents in the high quality group tended to have more time between high school and enlisting in the Army. This could indicate that the high quality white respondents first tried to work or further their education after high school and decided to join the Army to get help with these ambitions. This theory is somewhat reinforced by the second variable which contributes to high quality classification for white respondents. Factor 4_W indicates that the high quality white respondents surveyed joined the Army to get money for college or vo-tech school and to travel or get away from home to decide their future life plans. All of these reasons from Factor 4_W could be attributed to a person who tried other plans following high school and later considered to Army as a means to accomplish these previous goals.

C. DISCRIMINANT vs LOGIST MODEL

Based on the results presented in the analysis chapter of this thesis, it seems that the discriminant analysis model is less sensitive to unbalanced group membership of the data. This indicates that if the empirical distribution of the data is unknown or if it is believed to be skewed toward one particular group, then the discriminant

analysis model would be preferable. However, if this is not the case, logistic regression may be preferable due to the assumptions required by the discriminant analysis model. Further, the logistic regression model provides significance levels for model coefficients which are not computed during discriminant analysis. Ideally, both models should be used and the results compared as in this thesis to most accurately explore and model the data under observation.

APPENDIX A SELECTED FREQUENCY TABLES

TABLE 15 SEX REPORTED ON MEPRS/REQUEST

Sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Nc Match	72	.	.	.
Male	5233	90.4	5233	90.4
Female	558	9.6	5791	100.0

TABLE 16 MARITAL STATUS

Marital Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	5	.	.	.
No Match	72	.	.	.
Single	5159	89.2	5159	89.2
Married	552	9.5	5711	98.7
Separated	3	0.1	5714	98.8
Divorced	71	1.2	5785	100.0
Annulled	1	0.0	5786	100.0

TABLE 17 EDUCATION CERTIFICATION

Education Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	1	.	.	.
No Match	72	.	.	.
< H.S.	120	2.1	120	2.1
Certificate	76	1.3	196	3.4
GED	382	6.6	578	10.0
H.S.	4	0.1	582	10.1
HSDG	5023	86.8	5605	96.8
GED	50	0.9	5655	97.7
AA	52	0.9	5707	97.7
BA/BS & up	83	1.4	5790	100.0

TABLE 18 TERM OF ENLISTMENT

Enlistment Term	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	5	.	.	.
No Match	72	.	.	.
2-Year	384	6.6	384	6.6
3-Year	999	17.3	1383	23.9
4-Year	3765	65.1	5148	89.0
5-Year	417	7.2	5565	96.2
6-Year	217	3.8	5782	99.9
8-Year	4	0.1	5786	100.0

TABLE 19 AGE AT TIME OF ACCESSION

Age at Accession	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No Match	72	.	.	.
17 Years	521	9.0	521	9.0
18 Years	2127	36.7	2648	45.7
19 Years	1227	21.2	3875	66.9
20 Years	636	11.0	4511	77.9
21 Years	353	6.1	4864	84.0
22 Years	244	4.2	5108	88.2
23 Years	189	3.3	5297	91.5
24 Years	111	1.9	5408	93.4
25 Years	86	1.5	5494	94.9
26 Years	63	1.1	5557	96.0
27 Years	63	1.1	5620	97.0
28 Years	50	0.9	5670	97.9
29 Years	34	0.6	5704	98.5
30 Years	23	0.4	5727	98.9
31 Years	16	0.3	5743	99.2
32 Years	21	0.4	5764	99.5
33 Years	12	0.2	5776	99.7
34 Years	9	0.2	5785	99.9
36 Years	3	0.1	5788	99.9
37 Years	2	0.0	5790	100.0
> 37 Years	1	0.0	5791	100.0

TABLE 20 CASH BONUS

Cash Bonus	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No Match	72	.	.	.
Not Received	5249	90.6	5249	90.6
Received	542	9.4	5791	100.0

TABLE 21 ACF ELIGIBILITY

Army College Fund	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No Match	72	.	.	.
Not Eligible	4903	84.7	4903	84.7
Eligible	888	15.3	5791	100.0

TABLE 22 SELF-REPORTED RACIAL GROUP

Race	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Multiple Response	3	.	.	.
No Response	150	.	.	.
Indian or Alaskan	96	1.7	96	1.7
Asian or Pacific	122	2.1	218	3.8
Black	1541	27.0	1759	30.8
White	3951	69.2	5710	100.0

TABLE 23 MENTAL TEST CATEGORY

Mental Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No Match	72	.	.	.
4C,5	1	0.0	1	0.0
4B	1	0.0	2	0.0
4A	517	8.9	519	9.0
3B	1713	29.6	2232	38.5
3A	1490	25.7	3722	64.3
2	1870	32.3	5592	96.6
1	139	3.4	5791	100.0

APPENDIX B 1988 NRS INCENTIVE QUESTIONS

The following questions are reprinted from the 1988/89 USAREC Survey Form [Ref. 14:pp.].

In the next series of questions, use the following scale to rate HOW IMPORTANT each of the reasons listed below was in your decision to ENLIST.

1 - Not at all Important
2 - Somewhat Important
3 - Very Important
4 - I would not have enlisted except for this reason

33. I enlisted because I was unemployed and couldn't find a job.

34. I enlisted to give myself a chance to be away from home on my own.

35. I enlisted because the military will give me a chance to better myself in life.

36. I enlisted because I want to travel and live in different places.

37. I enlisted to get away from a personal problem.

38. I enlisted because I want to serve my country.

39. I enlisted because I can earn more money than as a civilian.

40. I enlisted because it is a family tradition to serve.

41. I enlisted to prove that I can make it.

42. I enlisted to get trained in a skill that will help me get a civilian job when I get out.

43. I enlisted so I can get money for a college education.

44. I enlisted because I want to be a soldier.

45. I enlisted so I can get money for civilian vocational, technical, or business school education.

46. I enlisted for the physical training and challenge.

47. I enlisted to take time out before deciding what I really want to do.

48. I enlisted because men and women are treated as equals in the military.

49. I enlisted because the military experience is beneficial to both men and women soldiers.

50. I enlisted because I want leadership training.

51. I enlisted because I like the retirement benefits.

52. I enlisted because I want the fringe benefits (e.g., health/dental care, low prices in military stores).

53. I enlisted to become a better person.

54. I enlisted to work with sophisticated, high-tech equipment.

55. I enlisted to become self-reliant.

56. I enlisted to learn to be a responsible mature person.

57. I enlisted to obtain a better job than the one I had.

58. Below are some reasons that people join the military. The next two questions contain <u>very similar</u> sets of reasons. They differ only in a few of the responses. Please be careful in answering; try to answer each question without comparing it to the other one.

A. Which of <u>these reasons</u> is your MOST IMPORTANT REASON for enlisting? (Mark only one)

- I was unemployed.
- To be away from home on my own.
- I want to travel.
- To get away from a personal problem.
- To serve my country.
- Earn more money.
- Family tradition to serve.
- To prove that I can make it.

- To get trained in a skill.
- Money for a college education.

B. Which of these reasons is your MOST IMPORTANT REASON for enlisting? (Mark only one)

- I was unemployed.
- To be away from home on my own.
- Chance to better myself.
- To get away from a personal problem.
- To serve my country.
- Earn more money.
- Family tradition to serve.
- To prove that I can make it.
- To get trained in a skill.
- Money for a college education.

APPENDIX C FACTOR LOADINGS

TABLE 24 FACTOR LOADINGS ALL RACIAL GROUPS

Reason for Enlisting	Factor1	Factor2	Factor3	Factor4
Responsible Person	0.72	0.27	0.14	0.09
Self-Reliant	0.69	0.26	0.14	0.14
Better Individual	0.66	0.36	0.17	0.01
Better Myself	0.51	0.30	0.23	-0.00
Money for College	0.24	-0.06	0.12	0.21
Be a Soldier	0.23	0.67	-0.05	0.05
Serve My Country	0.20	0.64	-0.02	0.01
Leadership Training	0.39	0.49	0.18	0.08
Physical Training	0.36	0.46	0.05	0.20
Prove I Can Make It	0.30	0.34	0.16	0.30
Family Tradition	-0.03	0.31	0.05	0.22
Fringe Benefits	0.11	0.32	0.58	0.05
Retirement Benefits	0.05	0.42	0.53	-0.00
Get a Better Job	0.21	-0.03	0.51	0.08
Skill Training	0.22	-0.07	0.43	0.00
Earn More Money	0.05	0.06	0.41	0.20
Money for Vo-Tech	0.23	-0.05	0.29	0.20
Unemployment	-0.07	-0.01	0.23	0.22
Away From Home	0.16	0.16	0.09	0.43
Decide Life Plans	0.11	0.09	0.02	0.39
Personal Problem	-0.03	0.01	0.04	0.36
Travel	0.16	0.29	0.14	0.29

TABLE 25 FACTOR LOADINGS BLACK GROUP ONLY

Reason for Enlisting	Factor1	Factor2	Factor3	Factor4
Responsible Person	0.70	0.19	0.18	0.15
Self-Reliant	0.63	0.19	0.23	0.16
Better Individual	0.59	0.34	0.27	0.02
Better Myself	0.55	0.25	0.19	-0.09
Be a Soldier	0.19	0.70	0.03	0.03
Serve My Country	0.18	0.67	0.04	-0.06
Leadership Training	0.29	0.50	0.22	0.13
Physical Training	0.28	0.45	0.11	0.23
Travel	0.19	0.23	0.20	0.21
Fringe Benefits	0.08	0.21	0.57	0.15
Retirement Benefits	0.02	0.32	0.54	0.07
Get a Better Job	0.23	-0.03	0.44	0.05
Vo/Tech Money	0.06	0.09	0.44	0.08
Skill Training	0.21	0.01	0.44	0.00
Earn More Money	0.05	0.02	0.43	0.18
College Money	0.12	0.01	0.33	0.05
Decide Life Plans	0.12	0.07	0.13	0.46
Be Away from Home	0.20	0.02	0.11	0.43
Personal Problem	-0.02	-0.01	0.01	0.42
Unemployment	-0.02	-0.00	0.06	0.39
Family Tradition	-0.10	0.28	0.07	0.35
Prove I Can Make It	0.22	0.24	0.19	0.33

TABLE 26 FACTOR LOADINGS WHITE GROUP ONLY

Reason for Enlisting	Factor1	Factor2	Factor3	Factor4
Responsible Person	0.77	0.14	0.11	0.16
Better Individual	0.75	0.23	0.14	0.07
Self Reliant	0.72	0.14	0.14	0.18
Better Myself	0.54	0.24	0.23	0.07
Leadership Training	0.48	0.40	0.14	0.11
Physical Training	0.43	0.43	0.02	0.22
Want to be a Soldier	0.36	0.61	-0.08	0.02
Serve My Country	0.36	0.57	-0.03	0.01
Prove I Can Make It	0.35	0.36	0.14	0.24
Family Tradition	0.03	0.31	0.04	0.10
Fringe Benefits	0.17	0.33	0.56	0.01
Get a Better Job	0.20	-0.05	0.53	0.10
Retirement Benefits	0.13	0.45	0.49	-0.06
Skill Training	0.19	-0.12	0.45	0.07
Earn More Money	0.03	0.12	0.42	0.15
Unemployment	-0.04	0.01	0.30	0.08
College Money	0.13	-0.04	0.05	0.44
Vo/Tech Money	0.14	-0.05	0.24	0.39
Life Plans	0.11	0.11	-0.01	0.39
Be Away From Home	0.12	0.25	0.09	0.38
Travel	0.16	0.33	0.14	0.33
Personal Problem	-0.04	0.05	0.06	0.24

APPENDIX D CLASSIFICATION RESULTS (20% WITHHELD DATA)

TABLE 27 DISCRIMINANT MODEL (COMBINED)

Actual Group	**Classified As Group**	
	High	*Other*
High	**346 (84.80%)**	**62 (15.20%)**
Other	**154 (51.51%)**	**145 (48.49%)**

TABLE 28 LOGISTIC MODEL (COMBINED)

Actual Group	**Classified As Group**	
	High	*Other*
High	**361 (88.48%)**	**47 (11.52%)**
Other	**166 (55.52%)**	**133 (44.48%)**

TABLE 29 DISCRIMINANT MODEL (BLACK ONLY)

Actual Group	**Classified As Group**	
	High	*Other*
High	**33 (56.90%)**	**25 (43.10%)**
Other	**66 (49.62%)**	**67 (50.38%)**

TABLE 30 LOGISTIC MODEL (BLACK ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**4 (06.90%)**	**54 (93.10%)**
Other	**2 (01.50%)**	**131 (98.50%)**

TABLE 31 DISCRIMINANT MODEL (WHITE ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**226 (60.75%)**	**146 (39.25%)**
Other	**74 (46.25%)**	**86 (53.75%)**

TABLE 32 LOGISTIC MODEL (WHITE ONLY)

Actual Group	Classified As Group	
	High	*Other*
High	**367 (98.66%)**	**5 (01.34%)**
Other	**155 (96.88%)**	**5 (03.12%)**

TABLE 33 LOGISTIC MODEL (BLACK ONLY, p=0.325)

Actual Group	**Classified As Group**	
	High	*Other*
High	**32 (55.17%)**	**26 (44.83%)**
Other	**61 (45.86%)**	**72 (54.14%)**

TABLE 34 LOGISTIC MODEL (WHITE ONLY, p=0.685)

Actual Group	**Classified As Group**	
	High	*Other*
High	**230 (61.83%)**	**142 (38.17%)**
Other	**74 (46.25%)**	**86 (53.75%)**

LIST OF REFERENCES

1. Vuono, Gen. Carl E., "Today's U.S. Army: Trained and Ready In an Era of Change," *Army*, v. 39, pp. 12-32, October 1989.

2. U.S. Department of the Army. *Army Focus*. Washington, D.C.: Pentagon, November 1989.

3. Meyers, Janet, "Selling The Services," *Government Executive*, v. 21, pp. 48-49, July 1989.

4. Flocke, Jenelle L., "The Madison Avenue Brigade," *Soldiers*, v. 43, pp. 18-20, May 1988.

5. Data Recognition Corporation. *1988 USAREC New Recruit Survey Codebook for Active Army Respondents, Trimester Report, Trimester 3 January - May 1989.*

6. The Rand Corporation P-6701, *Enlistment Standards In The Army*, by David J. Armor, October 1981.

7. The Rand Corporation R-2874-MRAL, *Recruit Aptitudes and Army Job Performance: Setting Enlistment Standards for Infantrymen*, by David J. Armor, Richard L. Fernandez, Kathy Bers, and Donna Schwarzbach, September 1982.

8. Naval Postgraduate School. *An Analysis Of Enlistment Motivations For United States Army Reserve Recruits*, by Linda Gorman and George W. Thomas, November 1989.

9. SAS Institute Inc. *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: SAS Institute Inc., 1985. 956pp.

10. Bennett, Spencer, and David Bowers, *An Introduction to Multivariate Techniques for Social and Behavioral Sciences*, The MacMillan Press, LTD, 1976.

11. Afifi, A.A., and Virginia Clark, *Computer-Aided Multivariate Analysis*, Lifetime Learning Publications, 1984.

12. SAS Institute Inc. *SUGI Supplemental Library User's Guide 1983 Edition*. Cary, NC: SAS Institute Inc., 1983. 402 pp.

13. Aldrich, John H., and Forrest D. Nelson, *Linear Probability, Logit, and Probit Models*, Sage Publications, 1984.

14. Data Recognition Corporation. *1988 USAREC New Recruit Survey Technical Manual, Yearly.*

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 52 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	1
4.	Prof. George Thomas. Code 54Te Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5000	1
5.	Prof. Laura Johnson, Code 55Jo Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	1
6.	Prof. Linda Gorman, Code 54Go Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5000	1
7.	Chief of Naval Operations OP 81 Navy Department Washington, D.C. 20350	1
8.	Commandant of the Marine Corps Headquarters, U.S. Marine Corps Washington, D.C. 20380	1

9. Deputy Undersecretary of the Army for Operations Research 1
Room 2E261, Pentagon
Washington, D.C. 20310

10. Commander 1
U.S. Army TRADOC Analysis Center
Fort Leavenworth, Kansas 66027

11. Captain Willis A. Woods 1
665 South 9th Street
Wytheville, Virginia 24382

12. Major John Hershberger 1
United States Army Recruiting Command
USARCPAE-ARA
Fort Sheridan, Illinois 60037-6000

13. Captain Steve Galing 1
United States Army Recruiting Command
USARCPAE-ARA
Fort Sheridan, Illinois 60037-6000